

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF: Masahiko IKEDA

GAU:

SERIAL NO: New Application

EXAMINER:

FILED: Herewith

FOR: SPEECH RECOGNITION APPARATUS

REQUEST FOR PRIORITY

COMMISSIONER FOR PATENTS  
ALEXANDRIA, VIRGINIA 22313

SIR:

☐ Full benefit of the filing date of U.S. Application Serial Number , filed , is claimed pursuant to the provisions of 35 U.S.C. §120.

☐ Full benefit of the filing date(s) of U.S. Provisional Application(s) is claimed pursuant to the provisions of 35 U.S.C. §119(e): Application No. Date Filed

☒ Applicants claim any right to priority from any earlier filed applications to which they may be entitled pursuant to the provisions of 35 U.S.C. §119, as noted below.

In the matter of the above-identified application for patent, notice is hereby given that the applicants claim as priority:

COUNTRY

APPLICATION NUMBER

MONTH/DAY/YEAR

Japan

2002-360356

December 12, 2002

Certified copies of the corresponding Convention Application(s)

☒ are submitted herewith

☐ will be submitted prior to payment of the Final Fee

☐ were filed in prior application Serial No. filed

☐ were submitted to the International Bureau in PCT Application Number

Receipt of the certified copies by the International Bureau in a timely manner under PCT Rule 17.1(a) has been acknowledged as evidenced by the attached PCT/IB/304.

☐ (A) Application Serial No.(s) were filed in prior application Serial No. filed ; and

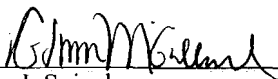
☐ (B) Application Serial No.(s)

☐ are submitted herewith

☐ will be submitted prior to payment of the Final Fee

Respectfully Submitted,

OBLON, SPIVAK, McCLELLAND,  
MAIER & NEUSTADT, P.C.

  
Marvin J. Spivak

Registration No. 24,913



22850

Tel. (703) 413-3000  
Fax. (703) 413-2220  
(OSMMN 05/03)

C. Irvin McClelland  
Registration Number 21,124

日本国特許庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日  
Date of Application:

2002年12月12日

出願番号  
Application Number:

特願2002-360356

[ST.10/C]:

[JP2002-360356]

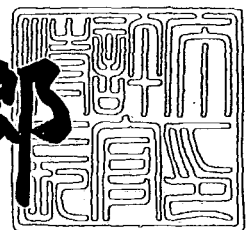
出願人  
Applicant(s):

三菱電機株式会社

2003年 1月14日

特許庁長官  
Commissioner,  
Japan Patent Office

太田信一郎



出証番号 出証特2002-3105498

【書類名】 特許願

【整理番号】 541028JP01

【提出日】 平成14年12月12日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 15/14

【発明者】

    【住所又は居所】 東京都千代田区丸の内二丁目2番3号 三菱電機株式会社  
社内

    【氏名】 池田 雅彦

【特許出願人】

    【識別番号】 000006013

    【氏名又は名称】 三菱電機株式会社

【代理人】

    【識別番号】 100089233

    【弁理士】

    【氏名又は名称】 吉田 茂明

【選任した代理人】

    【識別番号】 100088672

    【弁理士】

    【氏名又は名称】 吉竹 英俊

【選任した代理人】

    【識別番号】 100088845

    【弁理士】

    【氏名又は名称】 有田 貴弘

【手数料の表示】

    【予納台帳番号】 012852

    【納付金額】 21,000円

【提出物件の目録】

    【物件名】 明細書 1

特2002-360356

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声認識装置

【特許請求の範囲】

【請求項 1】 時系列に与えられる入力音声信号を特徴ベクトルに変換し、複数のフレームに区分して出力する音響処理部と、

予め準備された認識対象単語と音響モデルとに基づいて少なくとも 1 つの単語モデルを作成する単語モデル作成部と、

前記少なくとも 1 つの単語モデルと前記特徴ベクトルとの照合処理を、最大確率を与える状態系列に沿うことで最終累積確率を得るビタビアルゴリズムを用いて単語ごとに行う照合処理部と、

前記複数のフレームの各々に含まれる複数の状態について、確率に基づいて算出されるスコアの各フレーム中における最高値を記憶する最高値記憶部とを備え、

前記照合処理部は、

前記スコアの最高値に基づいて、前記複数の状態から、そのスコアを算出すべき計算対象状態を選択し、該計算対象状態以外の状態についてはスコアの算出を省略する間引き処理を行う、音声認識装置。

【請求項 2】 前記照合処理は、マトリックス状に配置された前記複数の状態に対して、それぞれが有する前記スコアを累積しつつ最終状態に到達するまでに取りうる複数のパスのうち、最大の累積スコアを与えるパスを特定することで、前記累積スコアを照合結果として取得する隠れマルコフモデルを用いた照合処理であって、

前記照合処理部の前記間引き処理は、

前記照合処理に際して、スコア算出の判断対象となっている現状態に至る前の前状態におけるスコアが、前記最高値記憶部に記憶された前記スコアの最高値に基づいて設定された所定の範囲内にある場合に、前記現状態を前記計算対象状態とし、前記前状態におけるスコアが前記所定の範囲外である場合は、前記現状態についてはそのスコアの算出を省略する処理を含む、請求項 1 記載の音声認識装置。

【請求項3】 前記照合処理部は、

前記最高値記憶部に記憶された前記スコアの最高値と、前記照合処理によって得られた各状態の最新スコアとをフレームごとに比較し、前記スコアの最高値を超える前記最新スコアが存在する場合には、前記スコアの最高値を前記最新スコアに書き換える機能をさらに含む、請求項2記載の音声認識装置。

【請求項4】 前記少なくとも1つの単語モデルは、複数の単語モデルであって、

前記音声認識装置は、

前記照合処理部から前記照合結果の情報を受け、最も最近に受けた最新単語モデルに対する前記照合結果と、既に受けている他の単語モデルに対する前記照合結果とを比較し、最も良好な最良照合結果を判定する照合結果判定部をさらに備え、

前記照合処理部は、

前記照合処理によって得られた各フレーム中の各状態の最新スコアの最高値を取得し、フレームごとに所定の一時記憶部に記憶される機能と、

前記照合結果判定部の判定結果の情報とを受け、前記最新単語モデルに対する前記照合結果が、前記最良照合結果である場合に、前記最高値記憶部に記憶された前記最新スコアの最高値を、前記一時記憶部に記憶させた前記各フレーム中の各状態の最高値に書き換える機能をさらに含む、請求項2記載の音声認識装置。

【請求項5】 前記少なくとも1つの単語モデルは、複数の単語モデルであって、

前記単語モデル作成部は、

前記複数の単語モデルを所定の共通項に基づいて複数の単語モデル集合に分類して出力する機能を備え、

前記音声認識装置は、

前記複数の単語モデル集合を受け、各単語モデル集合からそれぞれ代表となる代表モデルを選んで前記照合処理部に与え、前記代表モデルを用いた照合結果に基づいて前記単語モデル集合内の残りの単語モデルに前記照合処理を施すか否かを決定する照合対象単語選択部をさらに備える、請求項2記載の音声認識装置。

【請求項6】 前記単語モデル作成部は、  
前記認識対象単語のうち、先頭から数えて2つ以上で予め定めた個数の音の類似性を前記所定の共通項として用いて分類を行う、請求項5記載の音声認識装置。

【請求項7】 前記単語モデル作成部は、  
前記認識対象単語のうち、単語長を前記所定の共通項として用いて分類を行う、請求項5記載の音声認識装置。

【請求項8】 前記単語モデルの作成部は、  
前記認識対象単語のうち、パワーの変動情報に基づいて、無音部もしくは低パワー部の出現回数を前記所定の共通項として用いて分類を行う、請求項5記載の音声認識装置。

【請求項9】 前記音声認識装置は、  
前記照合処理部から前記照合結果の情報を受け、最も最近に受けた最新単語モデルに対する前記照合結果と、既に受けている他の単語モデルに対する前記照合結果とを比較し、最も良好な最良照合結果を呈する単語モデルを、入力単語に相当する単語データとして出力する照合結果判定部をさらに備え、

前記単語モデル作成部は、  
前記照合結果判定部が出力する前記単語データを受けて、統計処理を行い、出力回数の多い単語モデルが、前記照合対象単語選択部において優先的に選択されるように優先順位を付与する機能を備える、請求項5記載の音声認識装置。

【請求項10】 前記音声認識装置は、  
前記照合処理部から前記照合結果の情報を受け、最も最近に受けた最新単語モデルに対する前記照合結果と、既に受けている他の単語モデルに対する前記照合結果とを比較し、最も良好な最良照合結果を呈する単語モデルを、入力単語に相当する単語データとして出力する照合結果判定部と、

前記単語モデル作成部によって生成された前記単語モデルのデータを一時的に記憶するモデル辞書部と、をさらに備え、

前記照合対象単語選択部は、

前記照合結果判定部が出力する前記単語データを受けて、統計処理を行い、出

力回数の多い単語モデルを優先的に選択するように、前記モデル辞書部に記憶された前記単語モデルのデータの並び換えを行う機能を備える、請求項 5 記載の音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は音声認識装置に関し、特に単語音声照合処理を高速化した音声認識装置に関する。

【0002】

【従来の技術】

従来の音声認識方法の一例として、特許文献 1 に開示される方法が挙げられる。すなわち、特許文献 1 においては、隠れマルコフモデル (Hidden Markov Model) によるネットワークを状態と接点 (ノード) により表現し、このネットワーク上においてビタビ (Viterbi) アルゴリズムにより、各状態に生じる音声認識候補について、認識処理に必要な項目をすべて累積照合スコアと組にして伝播、処理することで累積照合スコアの計算量を減らし、記憶量も比較的小さくて済む音声認識方法が開示されている。

【0003】

【特許文献 1】

特開平 8-221090 号公報 (第 4 欄～第 8 欄、図 1)

【0004】

【発明が解決しようとする課題】

しかし、上記手法は、ビタビアルゴリズムを使用してフレーム同期で処理する音声認識を前提としており、技術の適用に制限があった。

【0005】

本発明は上記のような問題点を解消するためになされたもので、単語ごとに行う音声認識の照合処理においても、照合処理数を削減して処理速度を高速化することが可能な音声認識装置を提供することを目的とする。

【0006】



## 【課題を解決するための手段】

本発明に係る請求項 1 記載の音声認識装置は、時系列に与えられる入力音声信号を特徴ベクトルに変換し、複数のフレームに区分して出力する音響処理部と、予め準備された認識対象単語と音響モデルとに基づいて少なくとも 1 つの単語モデルを作成する単語モデル作成部と、前記少なくとも 1 つの単語モデルと前記特徴ベクトルとの照合処理を、最大確率を与える状態系列に沿うことで最終確率を得るビタビアルゴリズムを用いて単語ごとに行う照合処理部と、前記複数のフレームの各々に含まれる複数の状態について、確率に基づいて算出されるスコアの各フレーム中における最高値を記憶する最高値記憶部とを備え、前記照合処理部は、前記スコアの最高値に基づいて、前記複数の状態から、そのスコアを算出すべき計算対象状態を選択し、該計算対象状態以外の状態についてはスコアの算出を省略する間引き処理を行う。

【0007】

## 【発明の実施の形態】

## &lt;序論&gt;

発明の実施の形態の説明に先立って、単語音声照合に用いる隠れマルコフモデル (Hidden Markov Model: 以後 HMM と呼称) について説明する。

【0008】

図 1 は、4 つの状態を連結して構成される単語に対する HMM 照合処理を模式的に示す図である。ここで、状態とは音声言語の最小単位である音素 (phoneme) に相当する。なお、音素とは、一般的には母音や子音として知られている音がそれである。

【0009】

図 1 においては、横軸には時系列で入力された入力単語 (音声) を所定長さのフレーム単位ごとに区分した場合のフレーム数 ( $i$ ) を表し、縦軸には登録された単語の音素番号 ( $j$ ) を表し、マトリックスの格子点には○印を配しているが、各格子点には、入力単語のフレームごとに抽出した音響特徴量と、登録単語の各状態における照合確率の情報が示されている。なお、以下においては音素番号を状態番号と呼称し、マトリックスの格子点を音素片と呼称する。

## 【0010】

図1に示すHMM照合処理は、図に向かって左下隅の開始状態 $S(0, 0)$ から、右上隅の最終状態 $S(I, J)$ に至るまでの状態遷移系列を矢印で示しており、状態遷移系列が1つではないことを併せて示している。例えば、ある状態 $S(i, j)$ に着目した場合、状態 $S(i, j)$ に至るには、詳細図に示すように2つのパス $P1$ および $P2$ が存在する。すなわち、パス $P1$ は、状態 $S(i-1, j)$ からのパスであり同じ状態番号からの遷移（自己ループと呼称）であり、パス $P2$ は、状態 $S(i-1, j-1)$ からのパスであり異なる状態番号からの遷移である。

## 【0011】

ここで、状態 $S(i-1, j)$ に達するまでの確率の累積値（累積スコア）を $P(i-1, j)$ とした場合、パス $P1$ を通して状態 $S(i, j)$ に至る場合の確率 $w_{k1}$ は下記の数式（1）で表される。なお、開始状態 $S(0, 0)$ のスコアは初期値として与えられる値であり、例えば $P(0, 0) = 1$ となる。

## 【0012】

## 【数1】

$$w_{k1} = P(i-1, j) \times a\{(i-1, j), (i, j)\} \times b\{(i-1, j), (i, j), Y_i\} \dots (1)$$

## 【0013】

ここで、 $a\{(i-1, j), (i, j)\}$ は、状態 $S(i-1, j)$ から状態 $S(i, j)$ への遷移確率、 $b\{(i-1, j), (i, j), y_i\}$ は、状態 $S(i-1, j)$ から状態 $S(i, j)$ への遷移において、音声特徴ベクトル $Y_i$ が出現する確率である。

## 【0014】

また、状態 $S(i-1, j-1)$ に達するまでの累積スコアを $P(i-1, j-1)$ とした場合、パス $P2$ を通して状態 $S(i, j)$ に至る場合の確率 $w_{k2}$ は下記の数式（2）で表される。

## 【0015】

## 【数 2】

$$w_{k2} = P(i-1, j-1) \times a\{(i-1, j-1), (i, j)\} \times b\{(i-1, j-1), (i, j), Y_i\} \dots (2)$$

## 【0016】

ここで、 $a\{(i-1, j-1), (i, j)\}$  は、状態  $S(i-1, j-1)$  から状態  $S(i, j)$  への遷移確率、 $b\{(i-1, j-1), (i, j), y_i\}$  は、状態  $S(i-1, j-1)$  から状態  $S(i, j)$  への遷移において、音声特徴ベクトル  $Y_i$  が出現する確率である。

## 【0017】

上記数式 (1)、(2) から得られた確率  $w_{k1}$  および  $w_{k2}$  に基づいて、状態  $S(i, j)$  における累積スコア  $P(i, j)$  は下記の数式 (3) で与えられる。

## 【0018】

## 【数 3】

$$P(i, j) = \max(w_{k1}, w_{k2}) \dots (3)$$

## 【0019】

すなわち、パス  $P_1$  および  $P_2$  を通る場合に、それぞれ得られる確率  $w_{k1}$  および  $w_{k2}$  のうち、大きい方を状態  $S(i, j)$  での累積スコア  $P(i, j)$  とする。

## 【0020】

上記処理を最終フレームまで行い、最終状態  $S(I, J)$  における累積スコア  $P(I, J)$  が単語スコアとなる。

## 【0021】

なお、パス元が 1 つしかない状態については、当該パス元のスコアを算入することで自らのスコアを算出し、上記数式 (3) は用いない。

## 【0022】

なお、上記数式 (1) および (2) については、対数表記することで加算式になるので、得られる確率については累積スコアと呼称している。

## 【 0 0 2 3 】

なお、上述したHMM照合処理は、left-to-rightモデルとして周知のモデルである。

## 【 0 0 2 4 】

HMM照合処理は、開始状態から最終状態に至るまでの、ある状態遷移系列に沿って信号が出力される累積スコアの大小によって入力単語と登録単語との類似性を判断するものであり、複数の登録単語に対して上述したHMM照合処理を行い、単語スコアが最も大きな登録単語が、入力単語に最も類似するものと判断される。このように、最大確率を与える状態系列に沿って確率を求めるアルゴリズムをビタビ (Viterbi) アルゴリズムと呼称する。

## 【 0 0 2 5 】

## &lt; A . 実施の形態 1 &gt;

## &lt; A - 1 . 装置構成および動作 &gt;

本発明に係る音声認識装置の実施の形態 1 の構成および動作について、図 2 ～ 図 4 を用いて説明する。

## 【 0 0 2 6 】

## &lt; A - 1 - 1 . 装置全体の動作 &gt;

図 2 は実施の形態 1 の音声認識装置 1 0 0 の構成を示すブロック図である。図 2 に示すように、時系列で入力された音声入力 A 1 は、まず音声分析器 1 1 に与えられフレームごとに音響特徴量が抽出される。すなわち、音声分析器 1 1 においては、音声信号に、例えば L P C (Linear Predictive Coding 線形予測) 分析を行って音声のパワースペクトルを取得し、当該パワースペクトルから、声帯の振動を主たる発生源とする音源信号のスペクトルと、肺や顎、舌などの調音器官により形成される音響フィルタ (調音フィルタ) のスペクトルを分離し、調音フィルタの特性のみに関連する情報を音響特徴量として抽出する。なお、音響特徴量の抽出には、ケプストラム (Cepstrum) 分析が用いられ、また、ケプストラム分析で得られたケプストラム係数を人間の聴覚特性に基づいたメルケプストラム (Mel Cepstrum) 係数に変換する処理が施されることがあるが、これらの音響特徴量の抽出には公知の技術を用いれば良いので、これ以上の説明は省略する。

## 【 0 0 2 7 】

音声分析器 1 1 で音響特徴量を抽出した後、音声区間検出器 1 2 においてパワー（音の強さ）に基づいて音声区間を検出して、音響特徴量の時系列データとして入力音声特徴ベクトル V 1 を出力する。なお、音声分析器 1 1 および音声区間検出器 1 2 を含めて音響処理部と呼称する場合もある。

## 【 0 0 2 8 】

入力音声特徴ベクトル V 1 は時系列に単語照合処理器 2 に与えられ、登録単語との HMM 照合処理を施される。

## 【 0 0 2 9 】

ここで、HMM 照合処理を施すための照合対象となる単語を選択するまでの動作について、照合対象単語選択器 3、単語モデル生成器 4 および単語集合作成器 5 の動作に基づいて説明する。

## 【 0 0 3 0 】

例えば、EEPROM (Electrically Erasable Programable ROM) で構成される認識対象単語辞書 7 には、例えばテキスト形式でひらがな表記された複数の単語（登録単語）が登録されており、単語集合作成器 5 はその中から、例えば、先頭の数音を共通項とし、先頭の数音が似ているものどうしで集合を作るように動作する。この動作に際しては、ひらがな表記された登録単語を、音響モデル記憶部 6 に登録された確率分布をマトリックス状に配置して表現された音響モデル（HMM）に書き換え、音響モデルどうしで比較することで上述した集合を作成する。

## 【 0 0 3 1 】

すなわち、上述したように、音響モデルは確率分布を有しているので、先頭の数音について音響モデルどうしで確率分布を比較し合うことで、分布状態の類似性を判断し、類似する音響モデルで集合を作るようにすれば良い。

## 【 0 0 3 2 】

そして、単語モデル生成器 4 では、単語集合作成器 5 で作成した単語集合に対して、単語照合処理器 2 で照合できる形式の単語モデルの集合に変換する動作を行う。

## 【 0 0 3 3 】

ここで、単語集合の作成および音響モデルへの変換は、入力音声特徴ベクトル V 1 が入力されるごとに、毎回行っても良いし、認識対象単語辞書 7 が更新されたときに作成し、単語集合作成器 5 内にて集合情報を保持するようにしても良い。また、単語モデル生成器 4 内にて単語モデルの集合としてを保持してもよい。

## 【 0 0 3 4 】

なお、音声分析器 1 1、音声区間検出器 1 2、照合対象単語選択器 3、照合結果判定器 9、単語モデル生成器 4 および単語集合作成器 5 の動作は、プログラムを実行する CPU (Central Processing Unit) によって実現できる。

## 【 0 0 3 5 】

単語モデル生成器 4 によって生成された単語モデルの集合は、照合対象単語選択器 3 に与えられ、そのうちから照合対象となる 1 つの単語モデルが選択される。

## 【 0 0 3 6 】

照合対象単語選択器 3 によって選択された 1 つの単語モデルは、単語照合処理器 2 に与えられ、入力音声特徴ベクトル V 1、すなわち入力音声との照合処理が行われる。この照合処理が、先に説明した HMM を用いた処理である。

## 【 0 0 3 7 】

単語照合処理器 2 では、照合対象単語選択器 3 によって次々と選択される複数の単語モデルに対して HMM 照合処理を施し、各単語モデルの最終的な累積スコアである単語スコアを得る。なお、単語照合処理器 2 の動作は、単語モデル生成器 4 および単語集合作成器 5 を構成する前述の CPU で実現できるが、別途設けられた DSP (Digital Signal Processor) によっても実現できる。

## 【 0 0 3 8 】

そして、照合結果判定器 9 においては、単語照合処理器 2 から与えられる各単語モデルの単語スコアを記憶し、最も単語スコアの高い単語モデルを音声入力された単語に相当するものと判断し、当該単語モデルの出力単語データ D 1 を出力する。なお、照合結果判定器 9 は、照合結果に関する情報 D 2 を照合対象単語選択器 3 にフィードバックする機能を併せて有し、照合対象単語選択器 3 では、当

該情報D2に基づいて選択動作の効率化を図る。

【0039】

ここで、単語照合処理器2における照合処理および照合対象単語選択器3における選択動作について、最高値記憶バッファ8および照合結果判定器9の動作を含めて、それぞれ図3および図4に示すフローチャートを用いて説明する。なお、照合処理については図1に示すHMM照合処理を参照して説明する。

【0040】

#### ＜A-1-2. 単語照合処理器の動作＞

単語照合処理器2の動作について図3を用いて説明する。

照合処理が開始されると、まず、時系列に与えられる入力音声特徴ベクトルV1のフレーム番号0のフレーム( $i=0$ )を照合対象に定める(ステップS11)。そして、まず、単語モデルの状態番号0( $j=0$ )を指定する(ステップS12)ことで、照合対象が状態S(0, 0)となる。なお、最終フレーム番号はJであり、最終状態番号はIとする。

【0041】

次に、ステップS13において、照合対象が状態S(0, 0)であるか否かを判断し、状態S(0, 0)である場合はステップS15に進んでスコアの取得を行う(ステップS13)。

【0042】

一方、ステップS13において状態S(0, 0)以外の何れかの状態S( $i, j$ )と判断された場合は、ステップS14において、パス元が計算対象状態であるかについて判定を行う。

【0043】

この動作は、スコア取得対象としている現在の状態S( $i, j$ )の1つ前の状態、すなわちパス元のスコアが、単語照合処理器2に接続される最高値記憶バッファ8に記憶された、フレームごとのスコアの最高値に基づいて設定された所定の範囲内にあるか否かを判定する動作である。

【0044】

より具体的には、最高値記憶バッファ8には、入力音声特徴ベクトルV1の各

フレームごとに、スコアの最高値が記憶されている。この値は、過去に行った同一入力との照合処理の結果として得られた値であるが、以下に説明するように、照合処理ごとに更新可能な値である。なお、音声認識装置 1 0 0 において一番最初に照合処理を行う場合には、デフォルト値として、予め予想される所定の値が設定されるようにしておけば良い。

## 【 0 0 4 5 】

そして、当該スコアの最高値に対して、例えば所定のパーセンテージ以内の値というようにスコアの範囲を設定し、パス元のスコアが当該範囲内にあるか否かを判定する。

## 【 0 0 4 6 】

パス元のスコアが上記範囲内にある場合は、当該パス元のスコアを算入候補とし、数式 (3) に基づいて状態  $S(i, j)$  の累積スコアを取得する (ステップ S 1 5)。そして、スコアの取得後はステップ S 1 6 に進む。

## 【 0 0 4 7 】

なお、パス元が 1 つしかない状態については、当該パス元のスコアを算入することで自らのスコアを算出し、数式 (3) は用いない。

## 【 0 0 4 8 】

一方、パス元のスコアが上記範囲外であると判定された場合は、状態  $S(i, j)$  についてはスコアの計算を省略し、ステップ S 1 6 に進む。

## 【 0 0 4 9 】

ステップ S 1 6 では、現状の状態番号が最終番号 (J) に達しているか否かを判断し、最終番号に達していない場合には、状態番号を 1 つインクリメントし、ステップ S 1 4 以下を繰り返す。

## 【 0 0 5 0 】

また、最終状態番号に達している場合にはステップ S 1 7 に進み、1 つのフレームにおいて状態番号 0 から J までの状態に対して行った照合処理で得られた各状態でのスコアと、最高値記憶バッファ 8 に記憶されている現在照合対象となっているフレーム番号のフレームにおけるスコアの最高値とを比較し、より高いスコアが得られている場合には記憶されているスコアの最高値を、新たに得られた



より高いスコアに更新する。

【0051】

次に、ステップS18において、現状のフレーム番号が最終番号（I）に達しているか否かを判断し、最終番号に達していない場合には、フレーム番号を1つインクリメントし、ステップS12以下を繰り返す。

【0052】

上記動作は、例えば、フレーム番号0のフレームについて状態番号0からJまでの状態に対しての照合処理が終了した後は、フレーム番号1のフレームについて状態番号0からJまでの状態に対して照合処理を行うことを意味している。

【0053】

なお、最終フレーム番号に達している場合には、照合対象単語選択器3によって選択された1つの単語モデルに対する照合動作が終了する。

【0054】

このように、所定の閾値に基づいて、スコアの計算を省略する状態を設けるようにすることで、照合処理に要する時間を短縮することができる。なお、HMM照合処理においては、図1に示したように、最終状態S（I，J）に至るまでの状態遷移系列は、状態（0，0）を始点としてほぼ対角線に沿う経路を採ることが多く、極端に外れた経路を通る可能性は小さく、図1の配列における左上部の角部領域や、右下部の角部領域についてはスコアの算出は不要である場合が多く、スコアの計算を省略しても支障はない。

【0055】

なお、図1を用いて説明したように、最終状態S（I，J）における累積スコアが単語スコアとなり、上記ステップS11～S18の動作を、照合対象単語選択器3によって次々と選択される複数の単語モデルに対して施すことで、各単語モデルの単語スコアを得る。

【0056】

< A-1-3. 照合対象単語選択器の動作 >

照合対象単語選択器3は、単語モデル生成器4によって生成された単語モデルの集合から照合対象となる1つの単語モデルを選択すると説明したが、これは図

4 にステップ S 2 4 ～ S 2 6 で示す基本動作であり、この基本動作に先立って、ステップ S 2 1 ～ S 2 3 に示す前処理動作を行うことができる。

## 【 0 0 5 7 】

すなわち、照合対象単語選択器 3 は、単語モデル生成器 4 によって生成された単語モデルの集合を受けるが、この集合が 1 つではなく複数である場合、複数の集合にそれぞれ含まれる複数の単語モデルに対して照合処理を行うとなると、最終的な出力単語データ D 1 の出力までに長時間を有する可能性がある。

## 【 0 0 5 8 】

そこで、単語モデルの集合が複数である場合は、各単語モデルの集合からそれぞれ代表モデルを選び、当該代表モデルを単語照合処理器 2 に与えて照合処理を施し、その結果得られた単語スコアについて、照合結果判定器 9 において予め設定された判定基準値との比較を行う。その結果、当該単語スコアが判定値からかけ離れた値である場合は、上記代表モデルを抽出した単語モデルの集合については照合処理を施すのに不適当な集合であると判断する動作が前処理動作である。

## 【 0 0 5 9 】

なお、照合処理を施すのに不適当であると判断された集合は照合対象から外されることになる。

## 【 0 0 6 0 】

上述した前処理動作を含めて、照合対象単語選択器 3 の動作について図 4 を用いてさらに説明する。

## 【 0 0 6 1 】

単語選択動作が開始されると、まず、ステップ S 2 0 において、単語モデル生成器 4 から入力された単語モデルの集合が複数であるか否かの判定を行い、複数である場合にはステップ S 2 1 に進み、単語モデルの集合が 1 つである場合はステップ S 2 4 に進む。

## 【 0 0 6 2 】

ステップ S 2 1 においては、単語モデル生成器 4 から入力された単語モデルの複数の集合から、それぞれ代表モデルを選択する。すなわち、単語集合作成器 5 の動作において説明したように、単語モデルの集合の作成においては、例えば、

先頭の数音について音響モデルどうしで確率分布を比較し合うことで類似する音響モデルで集合を作るが、このとき、類似性の高低で集合内の音響モデルを分別し、類似性の高い音響モデルどうしを集めるようにし、この集合の最も中心にある音響モデルを代表モデルとすれば良い。

## 【 0 0 6 3 】

次に、ステップ S 2 2 において、複数の代表モデルのうちから何れか 1 つを選択して単語照合処理器 2 に与え、HMM照合処理を施す。なお、この場合の選択は無作為に行えば良い。

## 【 0 0 6 4 】

単語照合処理器 2 での HMM照合処理の結果として得られた単語スコアは照合結果判定器 9 に与えられ、予め設定された判定基準値と比較される。この判定基準値は経験値に基づいて設定すれば良く、例えば、過去に得られた単語スコアの平均値等を用いれば良い。そして、当該判定基準値を越えるか否かの判定結果を情報 D 2 として照合対象単語選択器 3 にフィードバックする。

## 【 0 0 6 5 】

次に、ステップ S 2 3 において、上記判定基準値を越えるか否かの判定結果に基づいて、上記代表モデルを抽出した単語モデルの集合について照合対象集合か否かを判断する。そして、照合処理を施すのに不適當な集合であると判断した場合には、当該集合を照合対象から外し、他の集合を選択し（ステップ S 2 8）、ステップ S 2 1 以下の動作を繰り返す。

## 【 0 0 6 6 】

また、ステップ S 2 3 において、照合処理を施すのに適當な集合であると判断した場合には、ステップ S 2 4 において、当該集合から 1 つの単語モデルを選択する。そして、単語照合処理器 2 に与え（ステップ S 2 5）、図 3 を用いて説明した手順で照合処理を行う。

## 【 0 0 6 7 】

なお、ステップ S 2 6 において、集合内に未処理の単語モデルが存在するか否かを判断し、未処理の単語モデルが存在する場合にはステップ S 2 4 以下の動作を繰り返し、集合内の全ての単語モデルが処理されている場合には、ステップ S

27において、未処理の集合が存在するか否かを判断し、未処理の集合が存在する場合にはステップS28において新たに集合を選択する。なお、全ての集合が処理されている場合には選択動作を終了する。

【0068】

<A-2. 特徴的作用効果>

以上説明したように音声認識装置100においては、単語照合処理器2でのHMM照合処理において、複数の状態のうち、照合対象となっている現状態に対するパス元（すなわち前状態）のスコアが、単語照合処理器2に接続される最高値記憶バッファ8に記憶された、フレームごとのスコアの最高値に基づいて設定された所定の範囲内にあるか否かを判定し、パス元のスコアが上記範囲内にある場合は、当該パス元のスコアを算入対象として累積スコアを取得するものとし、パス元のスコアが上記範囲外である場合には、照合対象の状態についてはスコアの計算を省略する。

【0069】

このように、単語ごとに行う音声認識の照合処理においても、いわゆるビームサーチ法と同様な間引き処理を行うことができ、1つの単語に対する照合処理に費やす時間を削減できる。

【0070】

また、単語集合作成器5によって類似する単語どうしで集合を作成し、照合対象単語選択器3によって、各単語モデルから代表モデルを選び、当該代表モデルを単語照合処理器2に与えて照合処理を施し、その結果得られた単語スコアに基づいて、上記代表モデルを抽出した単語モデルの集合に対して照合処理を施すか否かを判断する前処理動作を行うので、照合処理に費やす時間を大幅に削減して、より高速な処理が可能となる。

【0071】

<B. 実施の形態2>

<B-1. 装置構成および動作>

本発明に係る音声認識装置の実施の形態2の構成および動作について、図5～図7を用いて説明する。

【0072】

## &lt;B-1-1. 装置全体の動作&gt;

図5は実施の形態2の音声認識装置200の構成を示すブロック図である。なお、図5において、図2を用いて説明した音声認識装置100と同一の構成については同一の符号を付し、重複する説明は省略する。

【0073】

図5に示すように、入力音声特徴ベクトルV1は時系列に単語照合処理器24に与えられ、登録単語とのHMM照合処理を施される。単語照合処理器24は、基本的には図2に示す単語照合処理器2と同様の動作を行うが、最高値記憶バッファ8の他に一時記憶バッファ28が接続され、最高値記憶バッファ8に記憶されているスコアの最高値の更新手順に若干の相違を有している。なお、単語照合処理器24の動作の詳細については後述する。

【0074】

また、単語集合作成器25は認識対象単語辞書7の中から、例えば、先頭の数音が似ているものどうしで集合を作るように動作するが、このとき照合結果判定器9から出力される出力単語データD1を受けて統計処理を行い、出力回数の多い単語が、照合対象単語選択器3において優先的に選択されるように、当該単語を含む単語集合の優先順位を高く設定したり、当該単語の単語集合内での優先順位を高めるように優先順位を付与する機能を併せて備えている。

【0075】

## &lt;B-1-2. 単語照合処理器の動作&gt;

単語照合処理器24の動作について図6を用いて説明する。なお、図6において、ステップS31～S36までの動作は、図3を用いて説明したステップS11～S16までの動作と同じであり、重複する説明は省略する。

【0076】

ステップS36では、現状の状態番号が最終番号(J)に達しているか否かを判断し、最終番号に達していない場合には、状態番号を1つインクリメントし、ステップS34以下を繰り返す。また、最終状態番号に達している場合にはステップS37に進む。

## 【0077】

ステップS37では、ステップS34～S36を繰り返すことで取得した1つのフレームにおける状態番号0からJまでの各状態でのスコアのうち、最高値となるスコアを、一時記憶バッファ28に記憶させる。なお、この記憶は一時的なものであり、最高値記憶バッファ8に記憶されている各フレームの最高値のように、比較的長期に渡って保持されるものではなく、最高値記憶バッファ8とは異なるバッファを使用する。

## 【0078】

1つのフレームにおけるスコアの最高値を記録した後、ステップS38において、現状のフレーム番号が最終番号(I)に達しているか否かを判断し、最終番号に達していない場合には、フレーム番号を1つインクリメントし、ステップS32以下を繰り返す。

## 【0079】

また、最終状態番号に達している場合にはステップS39に進み、最終状態S(I, J)における累積スコアである単語スコアを照合結果判定器9に与える。

## 【0080】

照合結果判定器9では、過去に受け取った単語スコアと、単語照合処理器24から受け取った最新の単語スコアとを比較し、最新の単語スコアが、これまでの最高値となっている場合には、その情報を情報D3として単語照合処理器24にフィードバックする(ステップS40)。

## 【0081】

単語照合処理器24では、情報D3を受け、ステップS39で出力した単語スコアが最高値となっている場合には、一時記憶バッファ28に記憶した各フレームでのスコアの最高値を最高値記憶バッファ8に書き込むことで、最高値記憶バッファ8の記憶内容を更新する(ステップS41)。

## 【0082】

最高値記憶バッファ8の記憶内容を更新後は、照合対象単語選択器3によって選択された1つの単語モデルに対する照合動作が終了する。

## 【0083】

また、ステップ S 3 9 で出力した単語スコアが最高値となっていない場合には、最高値記憶バッファ 8 の記憶内容は更新されず、照合対象単語選択器 3 によって選択された 1 つの単語モデルに対する照合動作が終了する。

【 0 0 8 4 】

#### ＜ B - 2 . 特徴的作用効果 ＞

以上説明したように音声認識装置 2 0 0 においては、単語照合処理器 2 4 での HMM 照合処理において、照合対象の状態に対するパス元のスコアが、単語照合処理器 2 4 に接続される最高値記憶バッファ 8 に記憶された、フレームごとのスコアの最高値に基づいて設定された所定の範囲内にあるか否かを判定し、パス元のスコアが上記範囲内にある場合は、当該パス元のスコアを算入して累積スコアを取得するものとし、パス元のスコアが上記範囲外である場合には、照合対象の状態についてはスコアの計算を省略する。このように、単語ごとに行う音声認識の照合処理においても、いわゆるビームサーチ法と同様な間引き処理を行うことができ、1 つの単語に対する照合処理に費やす時間を削減できる。

【 0 0 8 5 】

また、単語照合処理器 2 4 では、各フレームにおける各状態でのスコアの最高値を一時記憶バッファ 2 8 に記憶させ、1 つの単語モデルに対する照合処理が終了した後、当該単語モデルの単語スコアが最高値である場合にのみ、一時記憶バッファ 2 8 に記憶した各フレームでのスコアの最高値を最高値記憶バッファ 8 に書き込むことで、最高値記憶バッファ 8 の記憶内容を更新するので、例えば、一部のフレームだけで、たまたま照合結果が良好であるような単語モデルのスコアが最高値記憶バッファ 8 に記録されることで、不正確な照合結果が得られることが防止できる。

【 0 0 8 6 】

また、単語集合作成器 2 5 において類似する単語どうしで集合を作成し、照合対象単語選択器 3 によって、各単語モデルから代表モデルを選び、当該代表モデルを単語照合処理器 2 4 に与えて照合処理を施し、その結果得られた単語スコアに基づいて、上記代表モデルを抽出した単語モデルの集合に対して照合処理を施すか否かを判断する前処理動作を行うので、照合処理に費やす時間を大幅に削減

して、より高速な処理が可能となる。

【0087】

また、単語集合作成器25においては、類照合結果判定器9から出力される出力単語データD1を受けて統計処理を行い、出力回数の多い単語が、照合対象単語選択器3において単語集合の代表モデルになるように優先順位を付与するので、入力頻度の高い単語について優先的に照合対象にすることができ、例えば、音声入力される単語の語彙が少なく、しかも入力単語に偏りがある場合、照合的中率を飛躍的に高めることができ、照合処理速度をさらに高速化できる。

【0088】

< B - 3 . 変形例 >

以上説明した音声認識装置200の変形例の構成を図7に示す。なお、図7において、図2および図5を用いて説明した音声認識装置100および200と同一の構成については同一の符号を付し、重複する説明は省略する。

【0089】

図7に示す音声認識装置200Aにおいては、単語モデル生成器4によって生成された単語モデルの集合のデータは、モデル辞書バッファ27に与えられ、一時的に記憶される。

【0090】

そして、モデル辞書バッファ27に保持された単語モデルの集合のデータは、照合対象単語選択器23に与えられ、そのうちから照合対象となる1つの単語モデルが選択される。

【0091】

ここで、照合対象単語選択器23は、図2を用いて説明した照合対象単語選択器3と同様の機能を有しているが、照合結果判定器9から出力される出力単語データD1を受けて統計処理を行い、出力回数の多い単語が、照合対象単語選択器23において優先的に選択されるように、出力回数の多い単語を含む集合の照合順位を上げるようにモデル辞書バッファ27に保持された単語モデルの集合のデータの並べ換えを行う機能もさらに有している。なお、上記統計処理に基づいて、出力回数の多い単語の集合内での優先順位を高めるようにデータの並べ換えを



行うようにしても良い。

【0092】

このように、音声認識装置200Aにおいては、単語モデル生成器4によって生成された単語モデルの集合のデータを記憶するモデル辞書バッファ27を有し、照合対象単語選択器23においては、照合結果判定器9から出力される出力単語データD1を受けて統計処理を行い、出力回数の多い単語を優先的に選択するように、モデル辞書バッファ27に記憶された単語モデルの集合のデータの並べ換えを行うので、入力単語に偏りがある場合、照合の的中率を飛躍的に高めることができ、照合処理速度をさらに高速化できる。

【0093】

<C. 他の変形例>

以上説明した音声認識装置100および200の各々においては、単語集合作成器5または25が、先頭の数音が似ているものどうしで集合を作るように動作することを説明したが、これは一例であり、他には、登録単語の単語長で集合を作成するようにしても良い。

【0094】

すなわち、登録されている単語に基づいて作成された音響モデルは、音素と継続時間長に関する情報を有しており、単語長は容易に推定できるので、単語長に基づいて集合を作成することは容易である。

【0095】

この方式を採用する場合、音声入力された単語の単語長は、フレーム数と相関するので、フレーム数から入力単語長を推定し、照合対象単語選択器3において、当該入力単語長に近似する単語長を有する単語集合を優先的に選択して照合することで、さらに高速な照合処理が可能となる。

【0096】

また、音素の情報にはパワー（音の強さ）およびパワーの変動に関する情報も含まれているので、登録単語内のパワーの変動に基づいて、無音（もしくは低パワー）の回数に基づいて単語集合を作成しても良い。

【0097】

なお、単語の先頭の数音の類似性、単語長およびパワーの変動の何れを組み合わせて用いても良いことは言うまでもない。

【0098】

<D. 照合処理の他の例>

以上説明した実施の形態1および2においては、照合処理としてHMM照合処理を用いる例を示したが、DPマッチング法による照合処理を使用しても良い。以下にDPマッチング法について説明する。

【0099】

同じ人が同じ言語を発しても、その継続時間はその都度変わり、しかも非線形に伸縮する。このため、標準パターンと入力音声との比較においては、同じ音素どうしが対応するように、時間軸を非線形に伸縮する時間正規化を行う。

【0100】

ここで、対応付けるべき2つの時系列を $A = a_1, a_2, \dots, a_i, \dots, a_I$ と、 $B = b_1, b_2, \dots, b_j, \dots, b_I$ で表し、図8に示すように横軸を入力パターンフレームを時系列に並べた系列A、縦軸を標準パターンフレームを時系列に並べた系列Bとする平面を想定する。なお、標準パターンは複数種類準備されているので、その複数種類の標準パターンに対応した平面が複数枚想定される。この場合、A、B両系列の時間軸の対応関係、すなわち時間伸縮関数は、この平面上の格子点 $c = (i, j)$ の系列Fで表現される。

【0101】

そして、2つの特徴ベクトル $a_i$ と $b_i$ とのスペクトル距離を $d(c) = d(i, j)$ で表すと、系列Fに沿った距離の総和 $H(F)$ は下記の数式(4)で表される。

【0102】

【数4】

$$H(F) = \frac{\sum d(C_k) \cdot w_k}{\sum w_k} \dots (4)$$

【0103】

この総和 $H(F)$ の値が小さいほど系列Aと系列Bとの対応付けが良いことを

示す。

【0104】

ここで、 $w_k$ は系列Fに関連する正の重みである。これに、単調性と連続性、および極端な伸縮を防ぐための諸制限を加えることで、図9に模式的に示すような時間伸縮関数Fの制限、すなわち、パスに対する傾斜制限が与えられる。

【0105】

図9においては、横軸を入力音声のフレームとし、縦軸を辞書に記憶された単語のフレームとし、それぞれ、 $i$ 軸、 $j$ 軸としてDPマッチングのパスモデルの例を示している。

【0106】

図9に示すように、4つのパスP1.1、P1.2、P1.3およびP1.4を想定した場合、パスP1.3およびP1.4のように、辞書フレーム番号を変更することのないパスどうしが連続することは制限され、パスP1.4は計算対象から外される。なお、パスP1.1～P1.3は点 $(i, j)$ に集結している。

【0107】

図9のパスモデルの場合の累積計算を数式化したものが下記の数式(5)となる。

【0108】

【数5】

$$g(i, j) = \min \begin{pmatrix} g(i-1, j) \\ g(i-1, j-1) \\ g(i-1, j-2) \end{pmatrix} + d(i, j) \quad \dots(5)$$

【0109】

数式(5)において、 $g(i, j)$ は点 $(i, j)$ における累積距離、 $g(i-1, j)$ はパスP3の累積距離、 $g(i-1, j-1)$ はパスP2の累積距離、 $g(i-1, j-2)$ はパスP1の累積距離であり、 $d(i, j)$ は図示しない始点からのユークリッド距離である。

## 【0 1 1 0】

ここで、 $g(1, 1) = d(1, 1)$  とし、まず  $j = 1$  の場合に固定して  $i$  が  $I$  に達するまで、順次変化させながら上記数式 (5) を計算しする。そして、次に、 $j$  の値を 1 つインクリメントして  $i$  について再び同様に变化させて計算を行う。この動作を  $j = J$  に達するまで繰り返すことで、系列 A および系列 B の 2 つの時系列間での時間正規後の累積距離が得られる。

## 【0 1 1 1】

この累積距離が HMM 照合処理で説明した累積スコアに相当し、累積距離の大小によって入力単語と登録単語との類似性を判断することが、DP マッチング法による照合処理であり、本願発明において HMM 照合処理の代わりに DP マッチング法を使用することが可能である。

## 【0 1 1 2】

## 【発明の効果】

本発明に係る請求項 1 記載の音声認識装置によれば、照合処理部において、スコアの最高値に基づいて、複数の状態から、そのスコアを算出する計算対象状態を選択し、該計算対象状態以外の状態についてはスコアの算出を省略する間引き処理を行うので、単語ごとに行う音声認識の照合処理においても、いわゆるビームサーチ法と同様な間引き処理を行うことができ、1 つの単語に対する照合処理に費やす時間を削減できる。

## 【図面の簡単な説明】

【図 1】 HMM による照合処理を説明する概念図である。

【図 2】 本発明に係る実施の形態 1 の音声認識装置の構成を示すブロック図である。

【図 3】 本発明に係る実施の形態 1 の音声認識装置の動作を説明するフローチャートである。

【図 4】 本発明に係る実施の形態 1 の音声認識装置の動作を説明するフローチャートである。

【図 5】 本発明に係る実施の形態 2 の音声認識装置の構成を示すブロック図である。

【図 6】 本発明に係る実施の形態 2 の音声認識装置の動作を説明するフローチャートである。

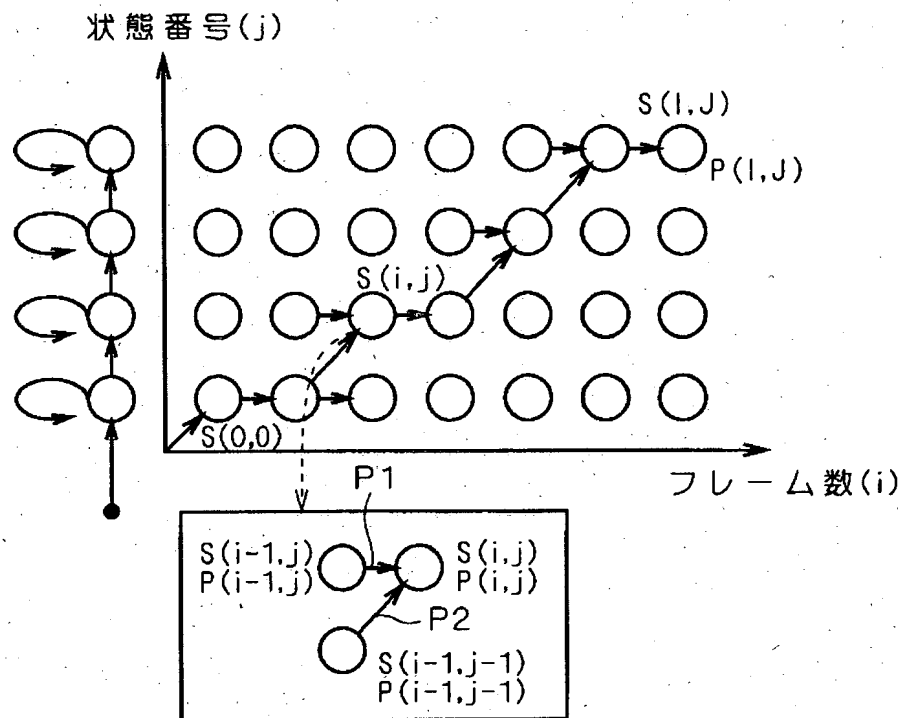
【図 7】 本発明に係る実施の形態 2 の音声認識装置の変形例の構成を示すブロック図である。

【図 8】 DP マッチング法による照合処理を説明する概念図である。

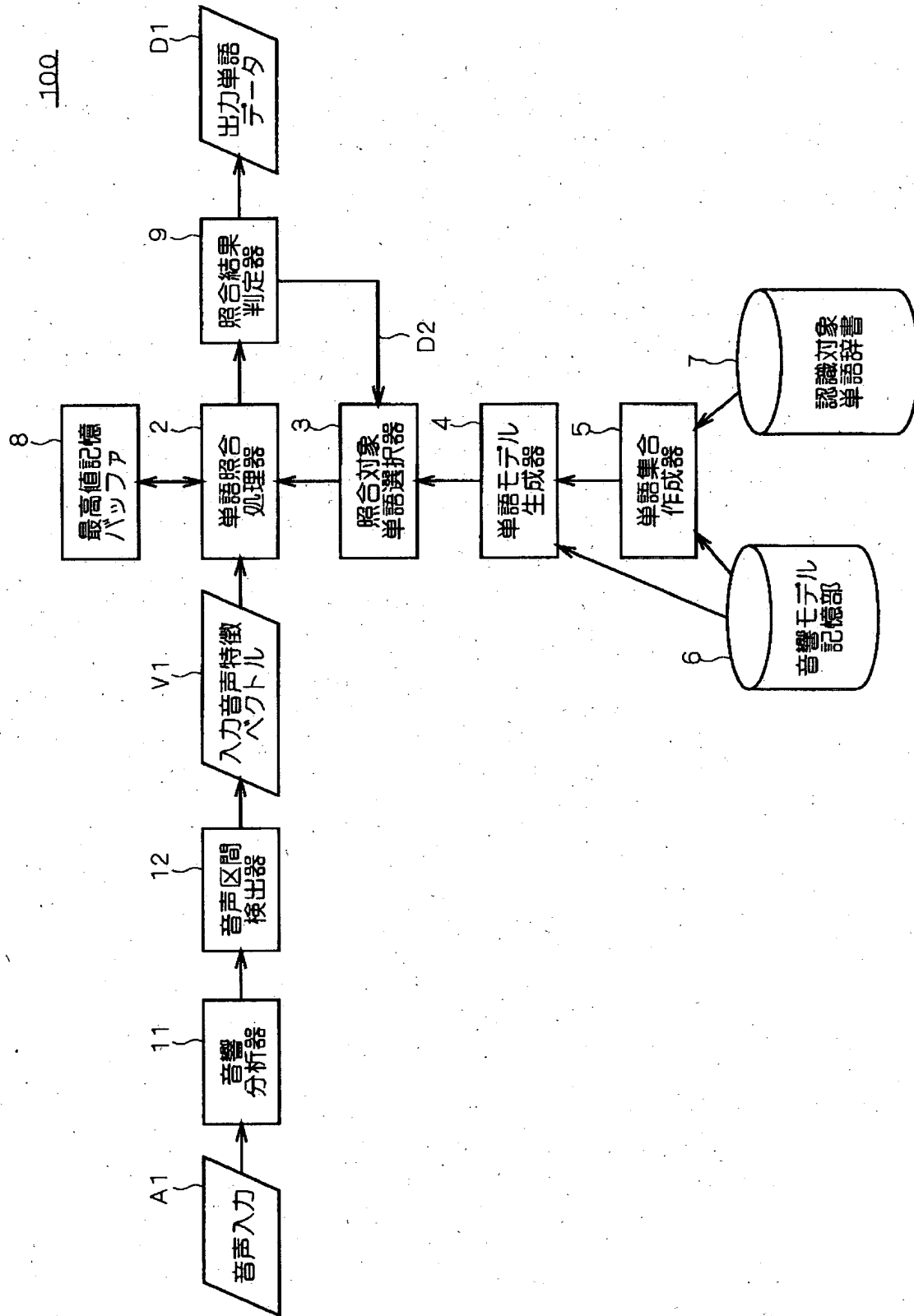
【図 9】 DP マッチング法による照合処理を説明する概念図である。

【書類名】 図面

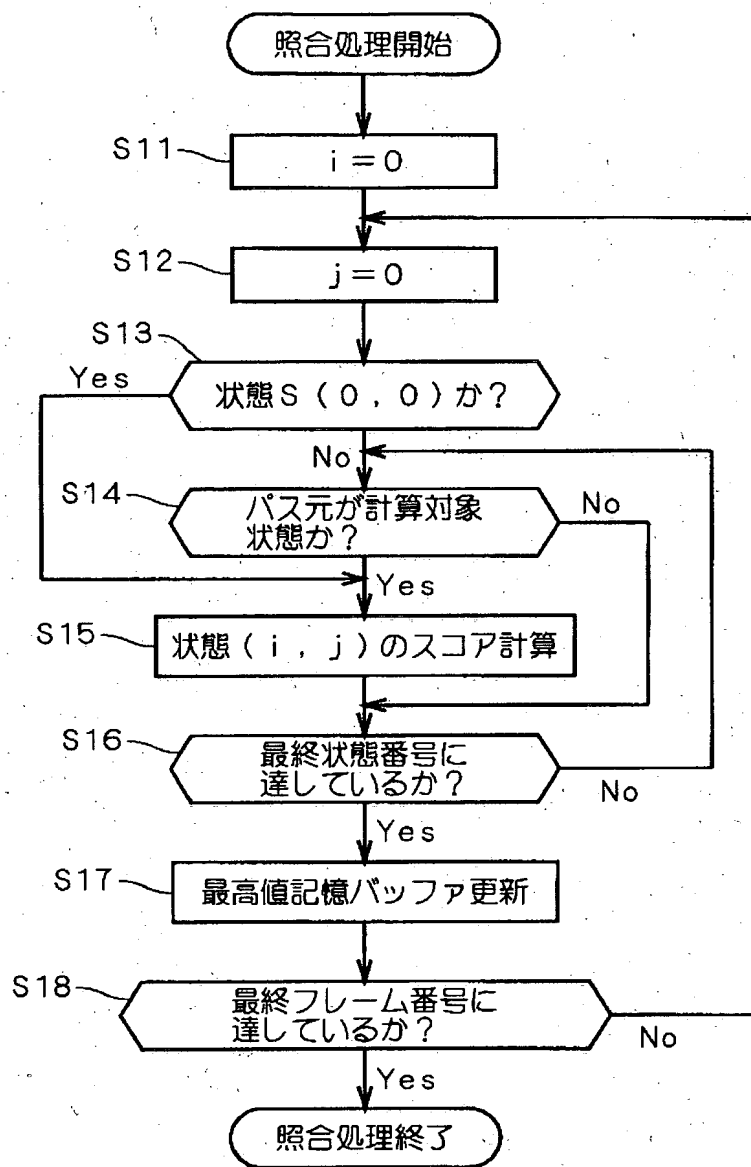
【図 1】



【図2】

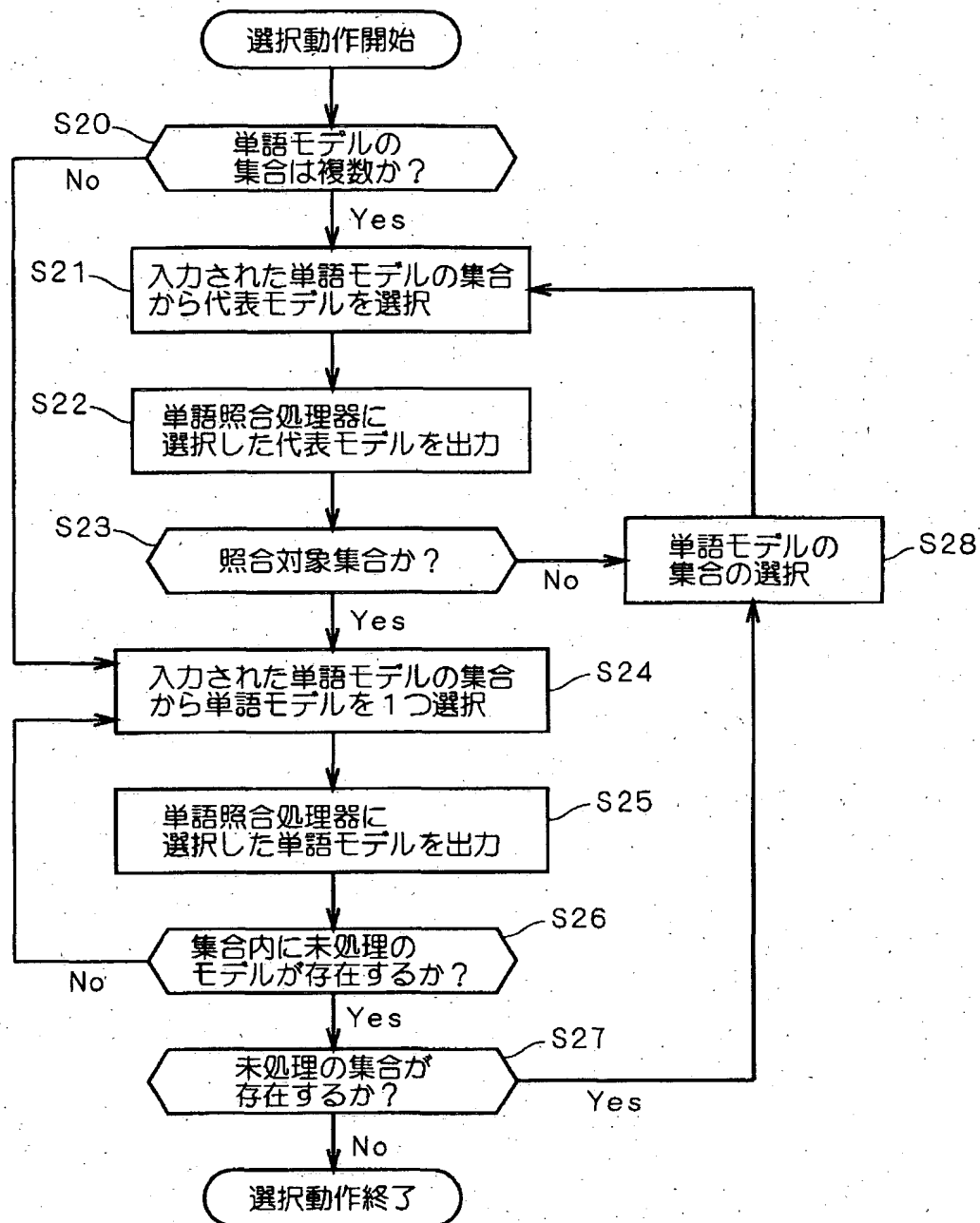


【図3】

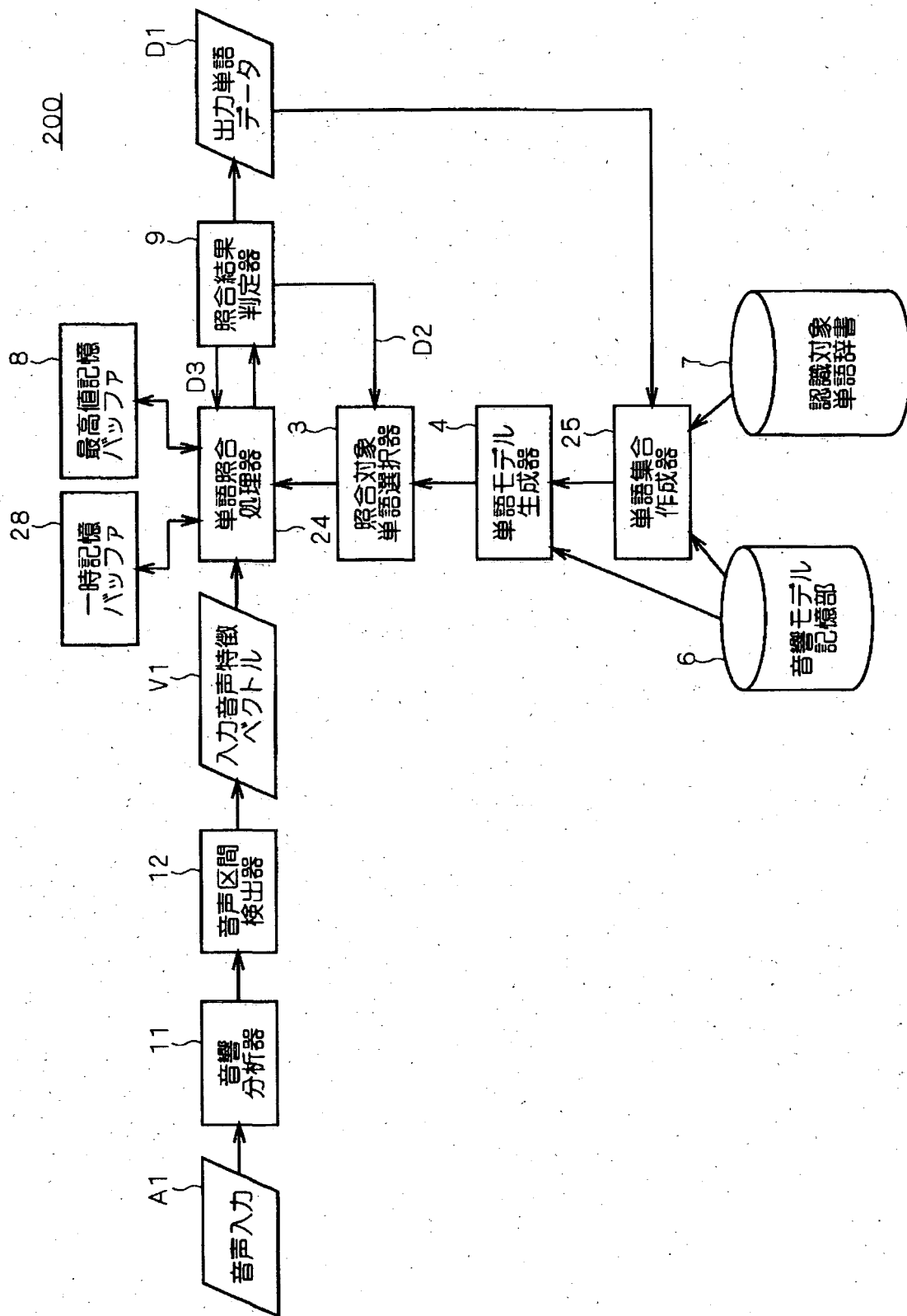




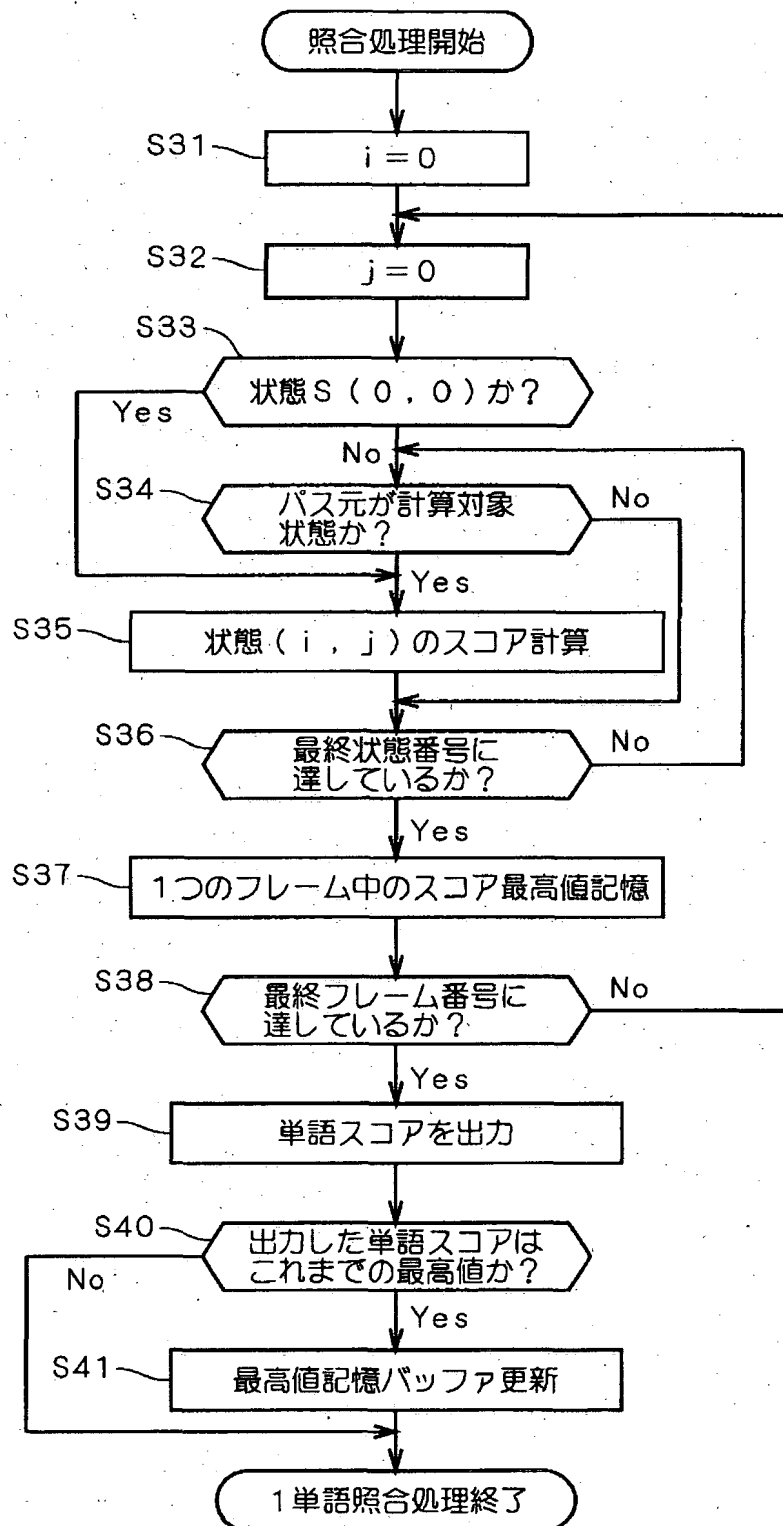
【図 4】



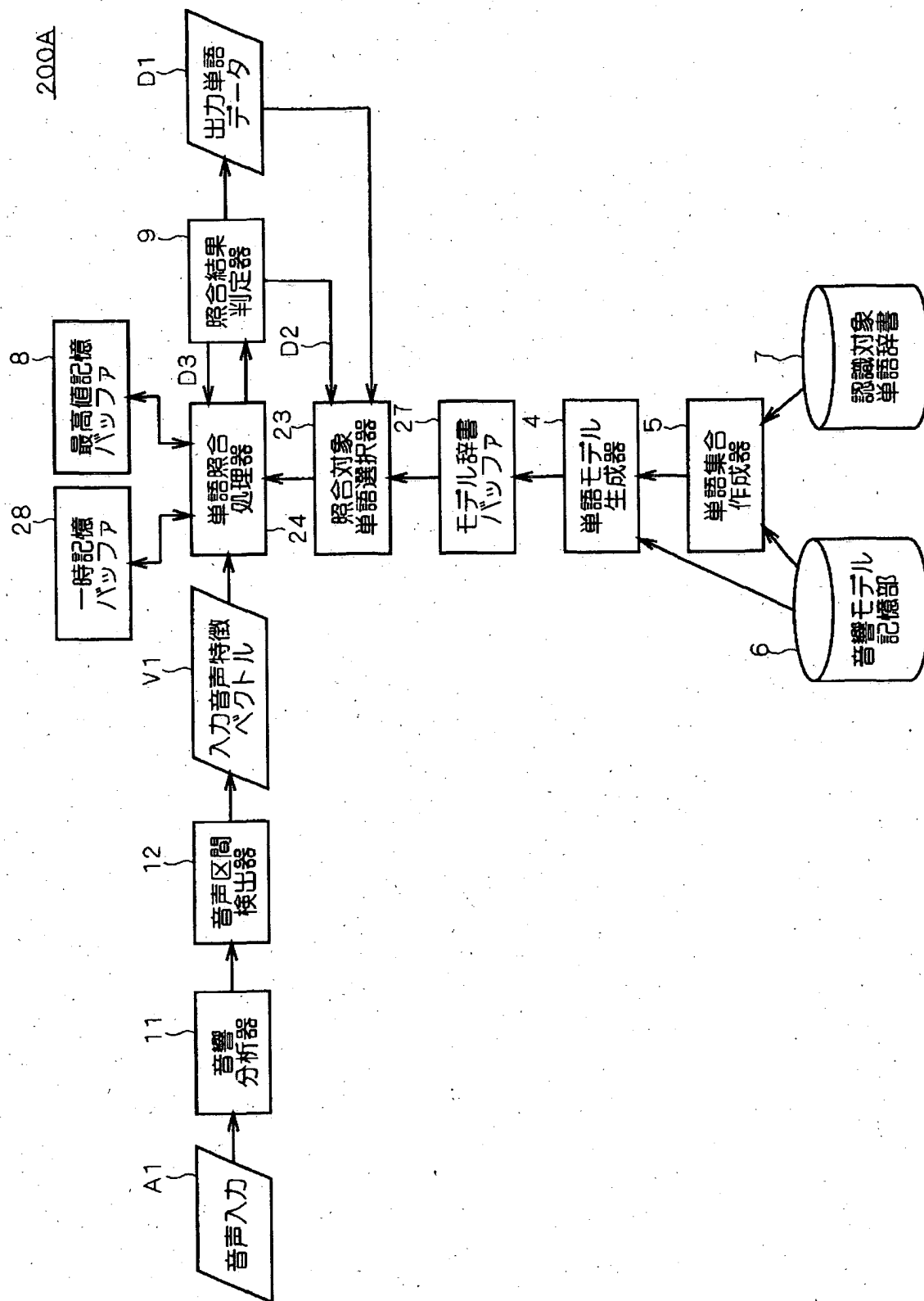
【図5】



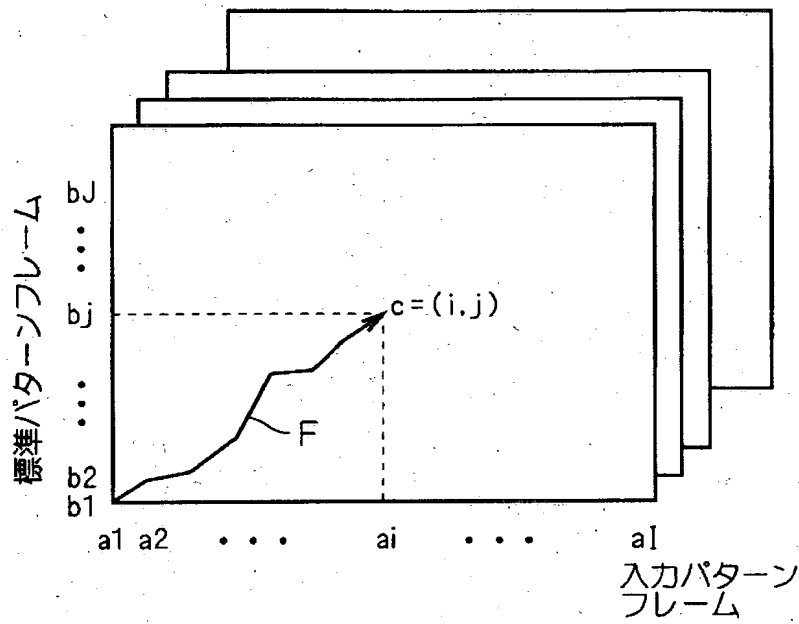
【図 6】



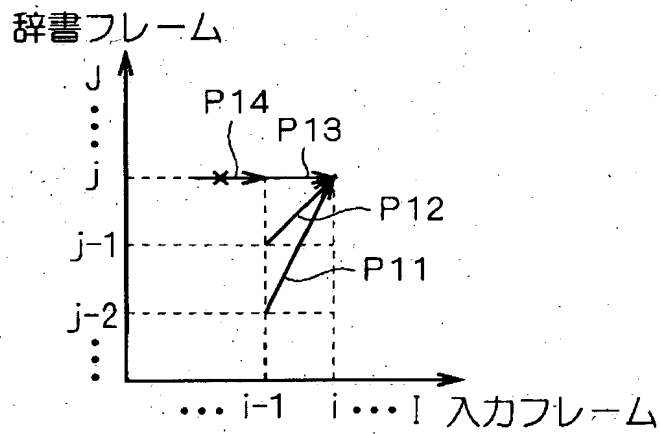
【図 7】



【図 8】



【図 9】



【書類名】            要約書

【要約】

【課題】    単語ごとに行う音声認識の照合処理においても、照合処理数を削減して処理速度を高速化することが可能な音声認識装置を提供する。

【解決手段】    単語モデル生成器 4 によって生成された単語モデルの集合は、照合対象単語選択器 3 に与えられ、そのうちから照合対象となる 1 つの単語モデルが選択される。単語照合処理器 2 では、照合対象となっている現状態に対するパス元のスコアが、単語照合処理器 2 に接続される最高値記憶バッファ 8 に記憶された、スコアの最高値に基づいて設定された所定の範囲内にあるか否かを判定し、パス元のスコアが上記範囲内にある場合は、当該パス元のスコアを算入対象として累積スコアを取得するものとし、パス元のスコアが上記範囲外である場合には、照合対象の状態についてはスコアの計算を省略する。

【選択図】            図 2

出 願 人 履 歴 情 報

識別番号 [000006013]

1. 変更年月日 1990年 8月24日  
[変更理由] 新規登録  
住 所 東京都千代田区丸の内2丁目2番3号  
氏 名 三菱電機株式会社